

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

Absolvování individuální odborné praxe
Individual Professional Practice in the
Company

2011

Michal Polcer

Prohlášení studenta

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně.

Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

Dne: 6. 5. 2011

.....

Podpis

Poděkování

Rád bych poděkoval vedoucímu bakalářské práce Ing. Tomáši Fabiánovi za odbornou pomoc a konzultaci při vytváření této práce. Dále společnosti Tieto Czech s.r.o. za možnost vykonání odborné praxe, hlavně mému konzultantovi Ing. Dušanu Šmelcerovi, který vedl mou práci během praxe ve společnosti.

Prohlášení zástupce spolupracující právnické nebo fyzické osoby

„Souhlasím se zveřejněním této bakalářské práce dle požadavků čl. 26, odst. 9 Studijního a zkušebního řádu pro studium v bakalářských programech VŠB-TU Ostrava.“

Dne: 6. 5. 2011

.....

Podpis zástupce

Abstrakt

Cílem této bakalářské práce bylo popsat open source nástroje dostupné pro oblast Business Intelligence a zjistit jejich možnost použití ve společnostech. První část práce se zabývá teoretickým popisem Business Intelligence. Cílem druhé části bylo najít dostupné open source nástroje, srovnat je a ohodnotit jejich vlastnosti. Ve třetí části bylo úkolem vybrat nejvhodnější nástroje pro nasazení do společností podle hodnocení a prakticky je pak vyzkoušet a otestovat jejich funkcionalitu. Závěrem práce mělo být doporučení, které nástroje je možné použít pro reálné řešení. Záměr práce se povedl naplnit.

Klíčová slova

Business Intelligence, Reporting, Dashboarding, Dolování dat, OLAP, ETL, SaaS.

Abstract

The aim of this thesis was to describe the open source tools available for Business Intelligence to determine their potential in companies. The first part deals with the theoretical description of Business Intelligence. The aim of the second part was to find available open source tools, compare them and evaluate their properties. The third task was to choose the most appropriate tools for use by companies under evaluation and then to try them practically and to test their functionality. The final part of my thesis should be a recommendation, which tools can be used for real solutions. The intention of work was successfully completed.

Key words

Business Intelligence, Reporting, Dashboarding, Data mining, OLAP, ETL, SaaS.

Seznam použitých symbolů a zkratk

BI	Business Intelligence
ETL	Extrakce, Transformace, Loading
OEM	Original Equipment Manufacture
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
SaaS	Software as a Service

Seznam tabulek

Tabulka 1: Bodové ohodnocení kritérií.....	7
Tabulka 2: Hodnocení nástrojů	13

Seznam obrázků

Obrázek 1: Architektura Pentaho BI suite.....	8
Obrázek 2: Architektura Jaspersoft	9
Obrázek 3: Architektura SpagoBI	10
Obrázek 4: Architektura Palo	11
Obrázek 5: Architektura OpenI	12
Obrázek 6: Magic Quadrant 2011	14

Obsah

1	Úvod	3
2	Úvod do Business Intelligence	4
2.1	Business Intelligence	4
2.2	Open source BI	4
2.3	Reporting	4
2.4	Dashboarding	5
2.5	Analýzy a vyhledávání	5
2.6	Distribuce, publikování a komunikace	5
2.7	ETL – Extrakce, transformace, load	5
2.8	OLAP	6
2.9	Data mining	6
2.10	SaaS – Software as a Service	6
3	Dostupné nástroje	7
3.1	Kritéria pro hodnocení nástrojů	7
3.2	Pentaho	8
3.3	Jaspersoft	9
3.4	SpagoBI	10
3.5	Palo Suite	11
3.6	OpenI	12
3.7	Hodnocení	13
4	Gartner Group	14
4.1	Magické čtverce – Magic Quadrant	14
4.2	Open source BI podle Gartner v roce 2010	15
4.3	Open source BI podle Gartner v roce 2011	15
4.3.1	Pentaho	15
4.3.2	Jaspersoft	16
5	Praktická část	18
5.1	Stažení a instalace	18
5.2	Testování funkcionality nástrojů	18
5.2.1	Pentaho	18
5.2.2	Jaspersoft	20

5.3	Zátěžový test – ETL nástroje.....	22
5.4	Hodnocení	22
6	Závěr.....	23
7	Literatura	24

1 Úvod

Odbornou praxi jsem vykonával ve firmě Tieto Czech s.r.o. v oddělení Business Intelligence (BI) v době od října 2010 do dubna 2011. Hlavním vedoucím mi byl Josef Mulka a konzultantem Dušan Šmelcer. Cílem mojí práce bylo zjistit, jaké open source nástroje jsou dostupné v oblasti BI, porovnat je, vybrané nástroje pak prakticky otestovat nad reálnými daty, analyzovat jejich funkcionalitu a možnost využití ve společnostech.

Práce se dělí do několika fází:

- Úvod do problematiky Business Intelligence (říjen, listopad)
- Nalezení dostupných open source nástrojů a jejich srovnání (prosinec, leden)
- Praktické testování nástrojů, které měly nejlepší hodnocení (leden, únor, březen)

Motivace

Firmy po celém světě mají společný cíl – chtějí být úspěšné. Aby se jim to povedlo, musí obstát v boji s konkurencí, mít co nejnížší náklady a naopak co nejvyšší zisk. Potřebují správně reagovat na výzvy trhu a rozhodovat se podle přesných informací. Jsou to informace o objednávkách, cenách, zboží, plánování a průběhu pracovního procesu nebo účetnictví, které jsou uloženy v databázích, v různých dokumentech či jiným způsobem. Správná analýza těchto dat pomáhá manažerům rozhodovat se o dalším chodu firmy. Může z ní vyplynout, který výrobek stáhnout z výroby, nebo na který se naopak zaměřit. Obecně řečeno umožňuje zvýšit efektivitu. Obor, který se tímto zabývá, se nazývá Business Intelligence.

Dříve panoval názor, že je Business Intelligence určen pouze pro velké společnosti, opak je ale pravdou. Postupem času se ukázalo, že se prosazuje i u středních a malých firem. V dnešní ekonomicky náročné době se objevují firmy, které mají požadavky na nízkonákladové BI řešení, protože si komerční řešení nemohou dovolit. Jelikož chtějí být ve firmě Tieto Czech s.r.o. na tuto situaci připraveni, bylo cílem mojí práce zjistit, které nástroje by bylo možné v budoucnosti využít.

Co se týče nabídky BI softwaru na trhu, je zde několik opravdu silných hráčů, kteří nabízejí velmi kvalitní produkty. Patří sem například Oracle, Microsoft, IBM, SAS, MicroStrategy, SAP a jiní. Jejich řešení jsou ale cenově náročná. Najdeme i volně dostupná řešení – tzv. open source BI. Mezi hlavní výrobce, které nabízejí komplexní řešení jako open source patří například Pentaho a Jaspersoft. Tato práce se zaměřuje právě na open source nástroje v oboru Business Intelligence.

2 Úvod do Business Intelligence

2.1 Business Intelligence

Business Intelligence (BI) jsou dovednosti, znalosti, technologie, aplikace, kvalita, rizika, bezpečnostní otázky a postupy používané v podnikání pro získání lepšího pochopení chování na trhu a obchodních souvislostech. Za tímto účelem provádí sběr, integraci, analýzu, interpretaci a prezentaci obchodních informací. Můžou zahrnovat samotné shromážděné informace nebo explicitní znalosti získané z informací.

Pojem se používá minimálně od roku 1958, s cílem podporovat lepší obchodní rozhodování. Tak lze také charakterizovat BI systémy jako systémy pro podporu rozhodování.

Business Intelligence aplikace poskytují historické, současné a prediktivní zobrazení obchodních operací, nejčastěji s využitím již získaných dat v datovém skladu nebo příležitostně přímo z provozních systémů. Běžné funkce BI aplikací zahrnují OLAP, reporting, podporu analýz, přehledové zobrazení (dashboarding), dolování dat, podnikové řízení výkonnosti a prediktivní analýzy.

BI aplikace zpracovávají data prodeje, výroby, financí a dalších zdrojů dat pro obchodní účely, především řízení výkonnosti podniku. BI aplikace mohou shromažďovat informace z různých částí společnosti a mohou porovnávat informace ve srovnatelných hodnotách.[1]

2.2 Open source BI

Open source má zejména ekonomický smysl, bourá striktní licenční politiku komerčních nástrojů a vede k minimálním nebo nulovým nákladům za licence. Open source licence zároveň umožňují rozšířit řešení na neomezený počet uživatelů a neomezený počet procesorů. Vždy existuje velká komunita vývojářů, která společně pracuje na nových vylepšeních systému. Cílem open source komunity je vývoj takového systému, který bude splňovat nejvyšší nároky na něj kladené. Velká komunita vývojářů je výborným zdrojem informací při řešení problémů se systémem. Vznikne tak plně funkční software bez dodatečné funkcionality – komerční nástroje obsahují velké množství funkcí, které pro běžný projekt nejsou vůbec potřeba. S open source systémem je možné si stáhnout pouze tu funkcionality, kterou uživatel aktuálně potřebuje, a postupně tuto funkcionality rozšiřovat.[2]

2.3 Reporting

Pro dobrého manažera je důležité být informován o dění v podniku. Jedním z hlavních úkolů BI je tyto informace poskytovat. V pojetí konceptu BI se v oblasti reportingu zdaleka nejedná o sadu reportů nebo sestav s tabulkami či grafy a o častém hledání, kontrole, sběru informací nebo manuálním zpracovávání nemůže být ani řeč. Reporting se stává integrovaným živým systémem reagujícím na to, kdo je přihlášený, jaké informace potřebuje a jaké oblasti a ukazatele sleduje, přičemž všechny informace jsou získávány z jednotného informačního zdroje, včetně stejné logiky uspořádání informací nebo výpočtu ukazatelů.[3]

2.4 Dashboarding

Dashboard (nástěnka) je jednou z nejoblíbenějších forem reportingu pro zprostředkování souhrnných informací. Zpravidla jej používají manažeři. Dashboard je anglický název pro přístrojovou desku u automobilu a vychází také z její jednoduchosti. Obsahuje pouze nejpodstatnější informace, zpravidla ve formě „budíků“ nebo kontrol. Na rozdíl od přístrojové desky v automobilu je ale počítačový dashboard velmi interaktivní a má široké možnosti grafických prvků pro zobrazení informací (tachometry, mapy, grafy, obrázky). Navíc může sloužit nejen k náhledu na sumární informace, ale i k simulacím typu co-když.[3]

2.5 Analýzy a vyhledávání

Bez analýz není poznání a nejsou ani klíčové podklady pro zlepšení. Analýza může být ve formě rychlého reportu nebo v sofistikovanější formě v podobě analýz časových řad, závislostí, odchylkových analýz nebo různých hitparád. Zajímavou formou analýzy je tzv. postup search & navigation, který má tři etapy. První je nalezení správných informací, což probíhá velmi podobně jako v internetovém vyhledávači typu Google. Druhou etapou je vlastní analýza vyhledaných informací. Třetí možnou etapou je pak sdílení výsledků například ve formě linku zaslaného do e-mailu.[3]

2.6 Distribuce, publikování a komunikace

Neméně důležitou částí Business Intelligence je schopnost informovat o dění v podniku za použití firemního intranetu, automatické distribuce reportů v podnikové síti nebo elektronické pošty. Standardem pro Business Intelligence je dnes již nabízet dynamický přístup k informacím přes internet. Celou řadu analýz, které se prováděly lokálně, lze provádět kdekoli a kdykoli je to třeba. Zajímavým trendem je také zobrazení a analýza firemních informací na mobilních zařízeních nebo tabletech typu iPad.[3]

2.7 ETL – Extrakce, transformace, load

ETL označuje právě mechanismus získávání dat z provozních systémů podniku (ekonomika, skladové hospodářství, výroba, odbyt atd.), jejich následné zpracování a poskytnutí aplikacím pro podporu rozhodování. Tato problematika je poměrně komplikovaná a představuje nezanedbatelný podíl (někdy až 70 %) v nákladech na budování systémů pro podporu rozhodování - nákladů ať finančních, tak i časových a lidských. V řadě projektů takovýchto systémů bývá bohužel podceňována, zvláště v prvotních fázích vývoje a budování. Tento přístup se potom může negativně promítnout do rychlosti vývoje a uvedení řešení do provozu, schopnosti flexibilně reagovat na měnící se podmínky, a v neposlední řadě do celkových nákladů řešení. Samotná zkratka ETL je tvořena počátečními písmeny třech slov - extrakce, transformace, load.

Extrakce znamená získání dat z primárních systémů. V typickém podniku je takovýchto systémů, zajišťujících životně důležité operace, celá řada. Zpravidla byly nasazovány v různých časových etapách a využívají různé, někdy i vysoce proprietární a málo otevřené technologie. Získat data z těchto systémů někdy může být nesnadný úkol, zvláště když si uvědomíme, že se nejedná o jednorázovou akci ale o periodickou činnost, jejichž cílem je poskytovat čerstvá data pro další zpracování v systémech datových skladů.

Transformace představuje proces zpracování dat získaných z primárních systémů do formy, odpovídající požadavkům systémů datových skladů. Transformace zahrnuje celou škálu operací od konverzí, matematických operací, filtrování, normalizace a denormalizace, až po sofistikované metody vytváření multidimenzionálních struktur. Data přicházející z primárních systémů mohou být (a ve valné většině případů jsou) "znečištěna" různými typy chybných či nekompletních údajů. Součástí transformací proto bývá také mechanismus kontroly kvality dat a čištění dat. Výstupem transformačního procesu jsou korektní a konsolidovaná data, která mají maximální informační hodnotu.

Závěrečnou fází je load, neboli naplnění zpracovaných dat do cílového systému datového skladu. Zde jsou pro uložení dat často využívány specializované technologie, které předpokládají použití různých proprietárních mechanismů pro rychlé a optimální ukládání dat.[4]

2.8 OLAP

OLAP je technologie uložení dat v databázi, která umožňuje uspořádat velké objemy dat tak, aby byla data přístupná a srozumitelná uživatelům zabývajícím se analýzou obchodních trendů a výsledků (Business Intelligence). Způsob uložení dat se svým zaměřením liší od běžněji užívaného OLTP, kde je důraz kladen především na snadné a bezpečné ukládání změn v datech v konkurenčním (víceuživatelském) prostředí. Základní rozdíly mezi OLAP a OLTP vyplývají z rozdílného použití – u OLAP se jedná o jednorázově nahrávaná data, nad kterými jsou prováděny složité dotazy, u OLTP jsou data průběžně a často modifikována a přidávána a to obvykle mnoha uživateli zároveň.[5]

2.9 Data mining

Data mining (angl. dolování z dat či vytěžování dat) je analytická metodologie získávání netriviálních, skrytých a potenciálně užitečných informací z dat. Data mining se používá v komerční sféře (například v marketingu při rozhodování, které klienty oslovit dopisem s nabídkou produktu), ve vědeckém výzkumu (například při analýze genetické informace) i v jiných oblastech (například při monitorování aktivit na internetu s cílem odhalit činnost potenciálních škůdců a teroristů).[6]

2.10 SaaS – Software as a Service

SaaS je model nasazení softwaru, kdy dochází k hostingu aplikace provozovatelem služby. Služba je dále nabízena zákazníkům přes internet. Eliminováním potřeb instalace a provozu aplikace na vlastních zařízeních se SaaS v poslední době stává oblíbeným způsobem provozu aplikace. SaaS vznikla jako reakce na potřebu snižování nákladů na software, rychlého nasazení a outsourcingu. Využíváním SaaS mohou firmy také redukovat přímé náklady na nákup softwaru, jelikož náklady na licenci on-demand bývají menší a zároveň není potřeba například licence na servery.[7]

3 Dostupné nástroje

Nejprve bylo nutné zjistit, jaké BI nástroje jsou dostupné na poli open source. Po průzkumu trhu jsem se rozhodl srovnávat tyto nástroje: Pentaho, Jaspersoft, SpagoBI, OpenI a Palo. Jedná se o produkty, které nabízejí komplexní BI řešení. Na trhu se také objevují nástroje, které se zaměřují pouze na některé funkce (například reporting), těmi jsem se však v mé práci nezabýval, protože původní požadavek byl na komplexní řešení. Všechny uvedené nástroje jsou vyvinuty v Javě, která je také nutná pro jejich instalaci (minimálně verze 1.5).

3.1 Kritéria pro hodnocení nástrojů

Před samotným porovnáním jsem stanovil kritéria, podle kterých jsem jednotlivé nástroje ohodnotil a určil maximální počet bodů, které mohou nástroje získat:

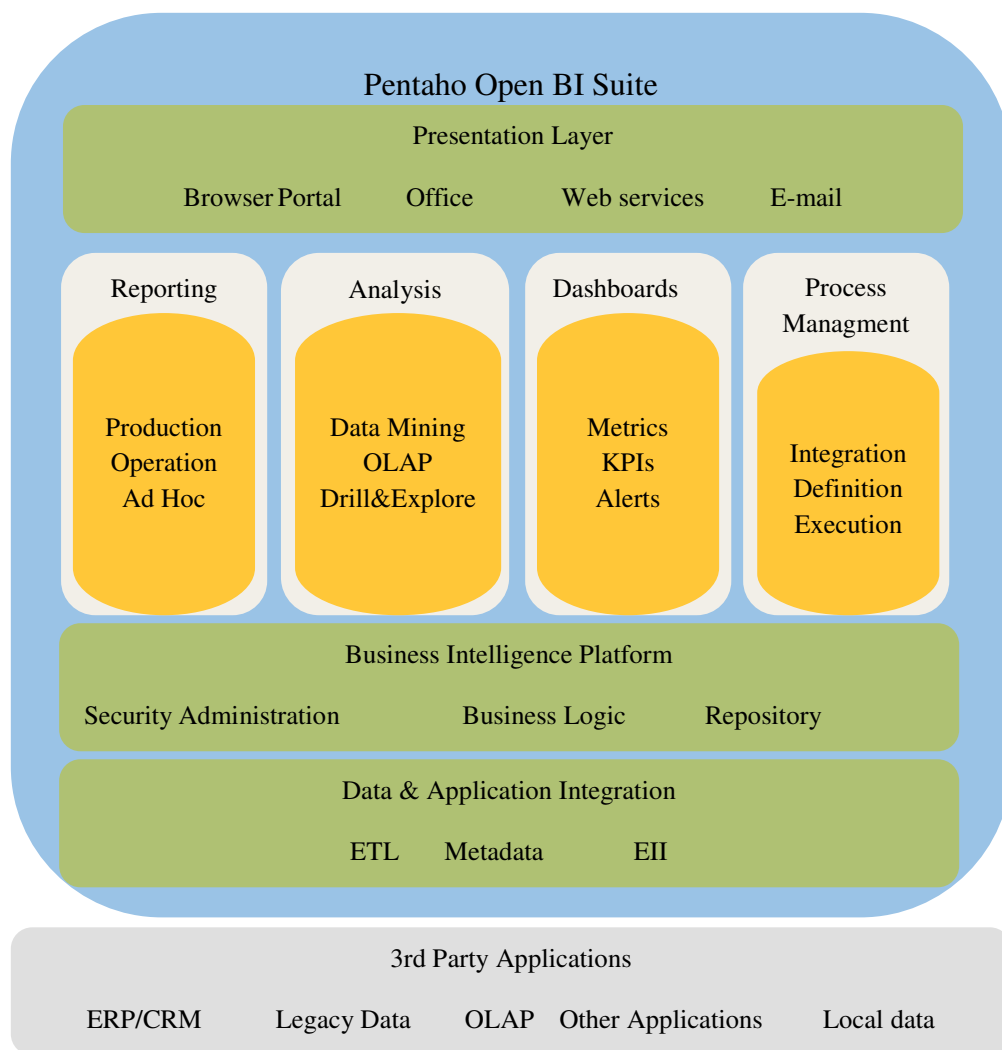
	Název kritéria	Maximální bodové ohodnocení
1.	Stažení softwaru, instalace	4b
2.	Komplexnost (vlastní řešení)	4b
3.	Ukázky praktického užití	2b
4.	Podporované operační systémy	3b
5.	Dostupná dokumentace	5b
6.	Kvalita podpory	4b
7.	Zkušební verze Enterprise edice	1b
8.	Podporované vstupní formáty dat	3b
9.	Výstupní formáty reportů	3b
10.	Software as a Service	1b

Tabulka 1: Bodové ohodnocení kritérií

U každého produktu je uveden stručnější popis, opírající se o získané informace, které jsou kompletně uvedeny v tabulce v externím souboru jako příloha, protože má větší rozměry, než stránka A4. Po zhodnocení všech nástrojů následuje tabulka s bodovými zisky ke všem určeným kritériím.

3.2 Pentaho

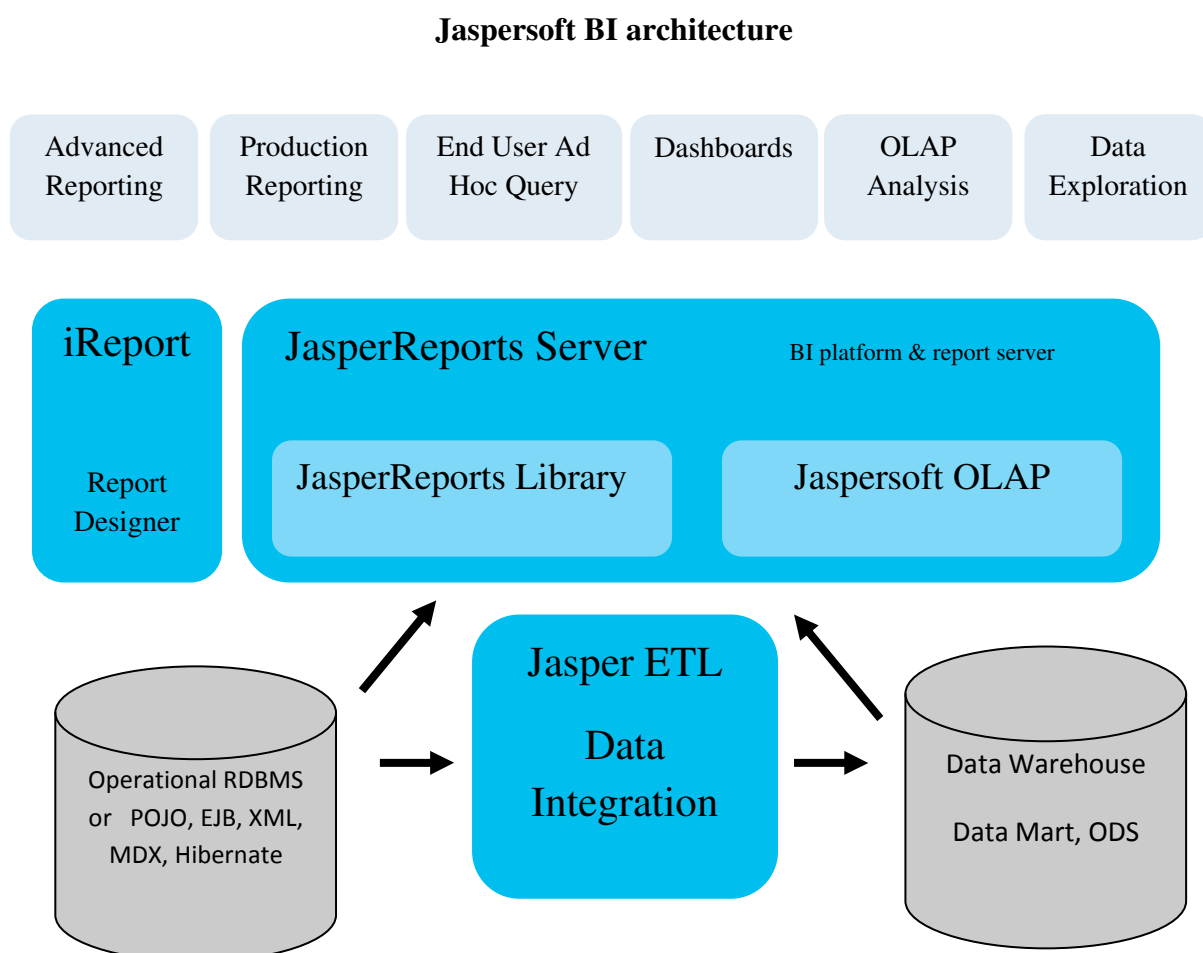
Společnost Pentaho nabízí pod licencí GNU LGPL produkt Pentaho BI suite verze 3.7, který je možno kompletně stáhnout z webových stránek. Produkt se skládá z několika nástrojů: Kettle pro datovou integraci, Mondrian pro OLAP analýzy, ReportDesigner pro tvorbu reportů, Weka pro data mining. BI platforma slučuje nástroje do aplikace, která je dostupná ve webovém prohlížeči. Pro stažení není nutná žádná registrace, instalace je popsána v dokumentaci a nedošlo při ní k žádnému problému. Na webových stránkách Pentaha jsou také dostupné různé tutoriály a dema, navíc je možno využít službu pro spuštění aplikace na jejich serveru a okamžitě si ji vyzkoušet. Pro stažení jsou dostupné verze pro operační systémy Windows, Linux i Mac. U každého produktu je velice důležitá dostupnost dokumentace. Pentaho ji poskytuje a je volně dostupná na webových stránkách pro každou součást Pentaho BI suite. Uživatelé mají možnost diskutovat pomocí fóra, a to jak uživatelského, tak developerského, mohou se přidat ke komunitě, využívat wiki stránky jednotlivých nástrojů. Najdeme zde i faq. Pentaho nám dává možnost 30 dní vyzkoušet Enterprise edici. U placené verze je také možnost využití SaaS (Software as a Service).



Obrázek 1: Architektura Pentaho BI suite

3.3 Jaspersoft

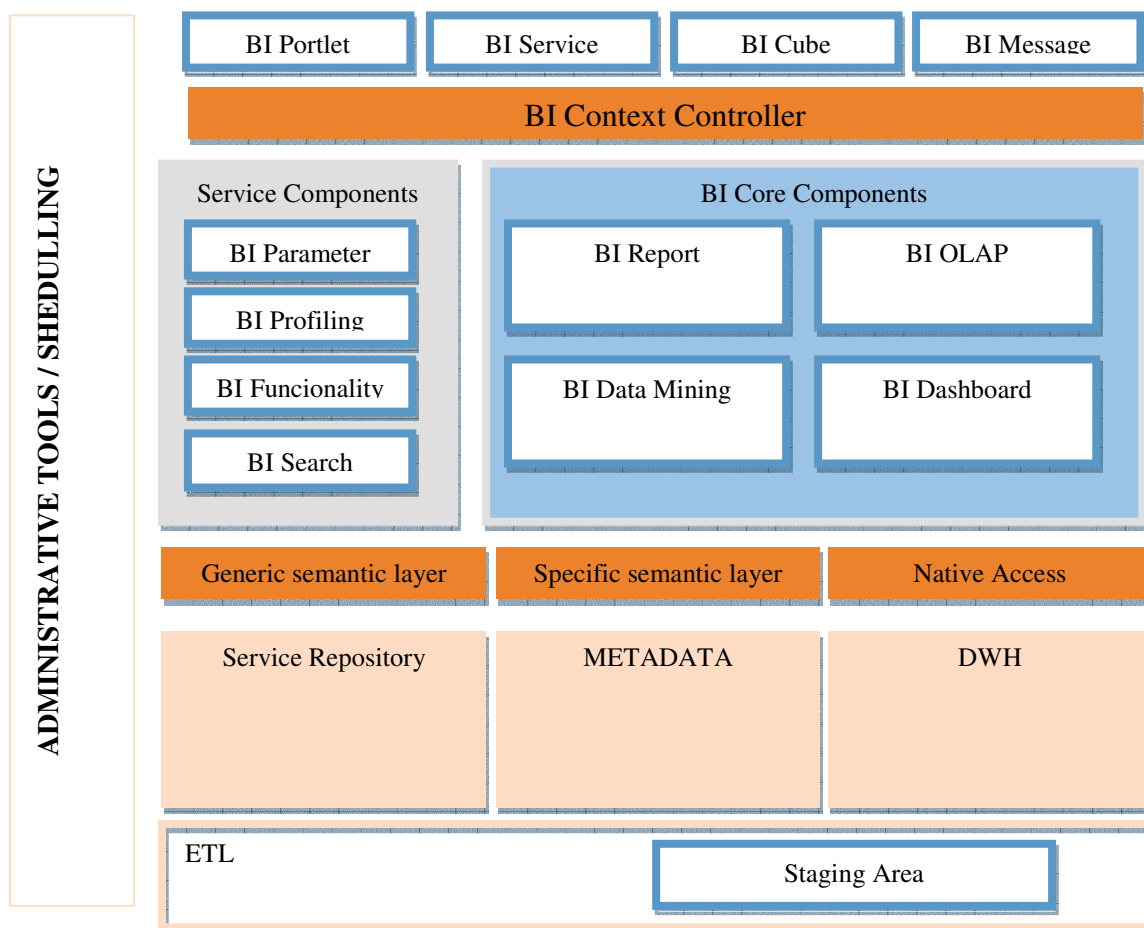
Společnost Jaspersoft nabízí pod licencí GPL produkt Jaspersoft BI suite verze 4. Pro stažení se musíme nejprve zaregistrovat, teprve potom si můžeme nástroj stáhnout. Instalace je popsána v instalační příručce. Produkt se skládá opět z několika nástrojů: JasperETL pro datovou integraci, JasperAnalysis pro OLAP analýzy, JasperReports pro tvorbu reportů. Schází zatím nástroj pro data mining. Kompletní platforma nese název JasperServer. Na webových stránkách jsou dostupné tutoriály a demo, nově je možno využít aplikaci běžící na jejich serveru a vyzkoušet práci s platformou. K dispozici jsou verze jak pro Windows, tak i pro Linux a v Enterprise edici také Mac a další. Menší problém nastává u dokumentace. Dostupné jsou základní uživatelské příručky, kompletní dokumentace už je placená a stojí 150 USD. Uživatelé mohou komunikovat na fóru, nebo využívat wiki stránky pro jednotlivé nástroje. Jaspersoft nabízí také možnost vyzkoušet 30 dní placenou verzi. U této verze poskytuje také službu SaaS.



Obrázek 2: Architektura Jaspersoft

3.4 SpagoBI

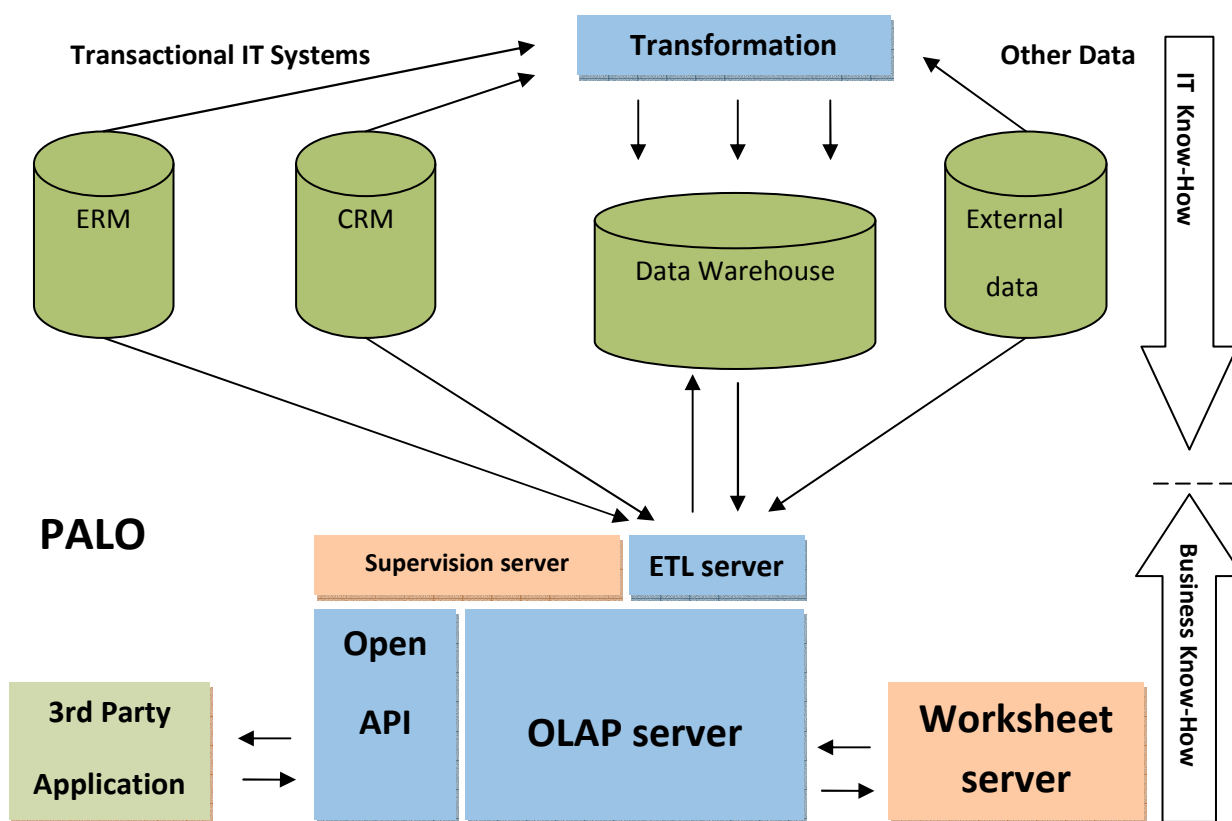
Společnost Engineering nabízí pod licencí GNU LGPL produkt SpagoBI verze 2.8. Ke stažení není nutná žádná registrace, nástroje jsou volně dostupné. Instalace je popsána v dokumentaci. Produkt se skládá z několika nástrojů: Talend OpenStudio pro datovou integraci, JPivot, JPalo nebo Mondrian pro OLAP analýzy, JasperReports a BIRT pro tvorbu reportů, Weka pro data mining. Tyto nástroje pak tvoří kompletní platformu SpagoBI Server. Na webových stránkách je dostupná ukázková webová aplikace a také tutoriály a dokumentace. Dále je možno využívat wiki stránky, faq nebo se připojit ke komunitě. Nechybí ani placená podpora. K dispozici jsou verze pro Windows a Linux.



Obrázek 3: Architektura SpagoBI

3.5 Palo Suite

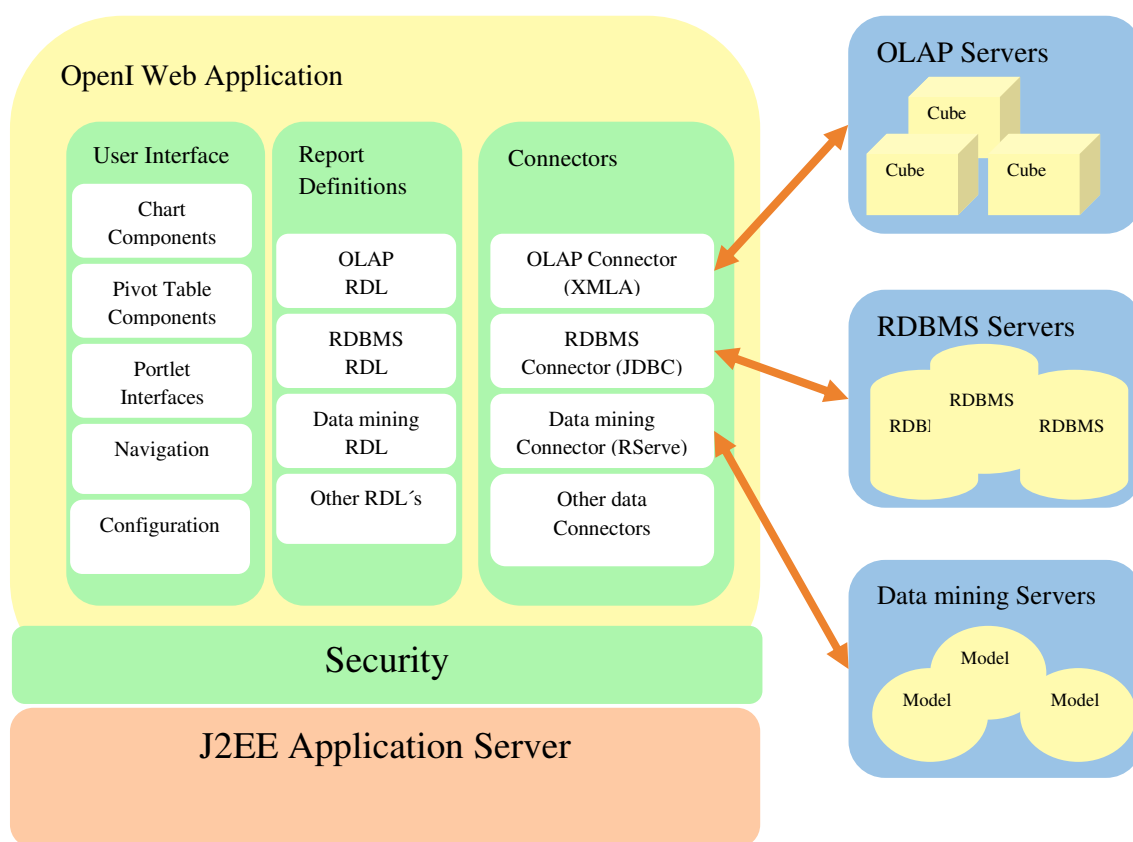
Společnost Jedox nabízí pod licencí GPL v2 produkt Palo Suite verze 3.2. Po registraci je možné stáhnout software z webových stránek. Po usilovném zkoušení se mi to bohužel nepodařilo, chyba byla pravděpodobně na straně poskytovatele produktu. Instalace je popsána v manuálu. Produkt se opět skládá z několika nástrojů: Palo ETL pro datovou integraci, Palo OLAP pro OLAP analýzy, Palo Spreadsheet pro tvorbu reportů, Palo for Excel pro práci s excelem, nástroj pro data mining neobsahuje. Všechny komponenty jsou zahrnuty ve webovém rozhraní Palo Web. Na webových stránkách jsou dostupná videa, online demo a také kompletní dokumentace. Stahovat lze verzi pro Windows a Linux. Uživatelé se mohou připojit ke komunitě, ke komunikaci využívat fórum, nebo podporu (telefonicky, emailem). Jedox nám dává možnost stáhnout si na 30 dní Enterprise edici. Poskytuje také službu SaaS.



Obrázek 4: Architektura Palo

3.6 OpenI

Pod licencí GPL v2 je dostupný produkt OpenI verze 2.0. Pro stažení není nutná žádná registrace. Tento produkt využívá některé již dříve zmíněné nástroje: jPivot nebo Mondrian pro OLAP analýzy, JaspersReports pro tvorbu reportů a R-project pro data mining. Nástroj pro datovou integraci není k dispozici. Instalace je popsána na wiki stránkách, kde lze také stáhnout uživatelskou příručku. Z domovské stránky lze spustit ukázkovou aplikaci a je možné si produkt vyzkoušet. Kromě wiki stránek je zde i komunita. OpenI je k dostání ve verzi pro Windows a Linux. Nabízí se i služba SaaS.



Obrázek 5: Architektura OpenI

3.7 Hodnocení

Podle mnou zvolených kritérií tak nejvíce bodů získalo Pentaho, jenž se ukázalo jako velmi komplexní a kvalitní nástroj. Na druhém místě ho následoval Jaspersoft, jenž by dosáhl téměř stejného výsledku, nebýt ztráty bodů kvůli placené dokumentaci. Jaspersoft byl dříve zaměřen hlavně na reporting a i když u něj stále schází nástroj pro data mining, jeho komplexnost je vyšší díky začlenění nástroje pro datovou integraci. Na dalších místech pak následují: 3. SpagoBI, 4. Palo a 5. OpenI.

Produkty na prvních dvou místech dosahují vyšší kvality, než ostatní srovnávané produkty, a poskytují funkcionalitu blížící se komerčním řešením, což také dokazuje rozbor podle společnosti Gartner Group. I proto jsem se je také rozhodl následně prakticky otestovat.

Kritérium	Pentaho	Jaspersoft	SpagoBI	Palo	OpenI
Stažení softwaru, instalace	4	3,5	3	2	3
Ukázky praktického užití	2	2	2	2	1
Podporované operační systémy	3	3	2	2	2
Dostupná dokumentace	5	3,5	4	4	2
Kvalita podpory	3	2	3	2	1,5
Zkušební verze Enterprise edice	1	1	0	1	0
Vstupní formáty dat	3	2,5	3	1,5	1
Výstupní formáty reportů	2	3	2	1	1
Software as a Service	1	1	0	1	1
Komplexnost produktu	4	3	4	3	3
Celkem	28	24,5	23	19,5	15,5

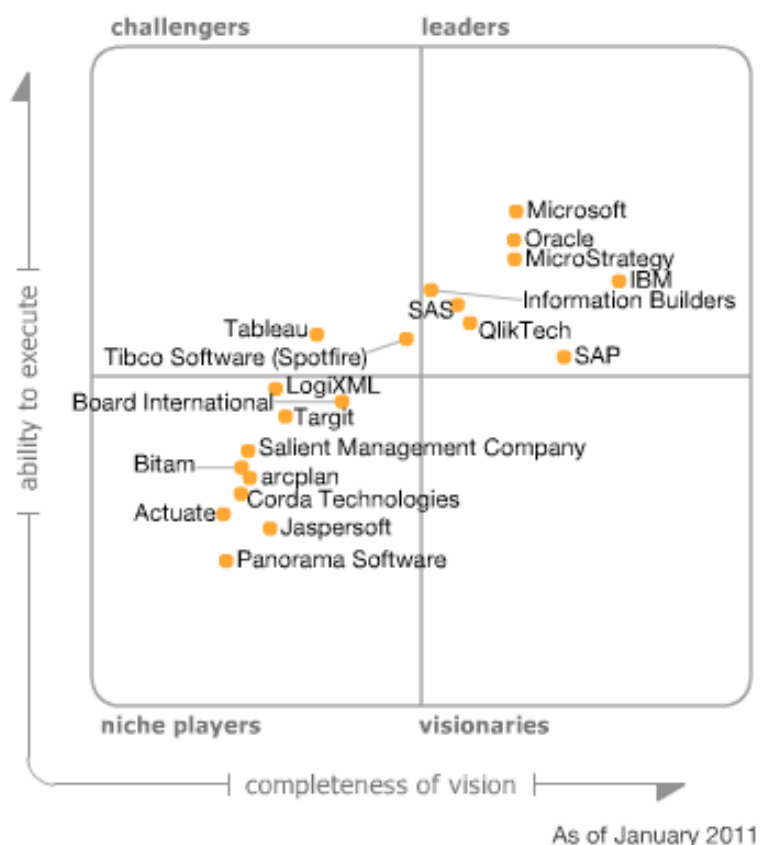
Tabulka 2: Hodnocení nástrojů

4 Gartner Group

Gartner je přední světovou společností zabývající se výzkumem informačních technologií. Podává technologické pohledy, které jsou nezbytné pro správné rozhodnutí při volbě technologie. Společnost byla založena v roce 1979 a sídlí ve Stamfordu v USA. Zaměstnává více než 4400 zaměstnanců, včetně 1200 výzkumných pracovníků, analytiků, konzultantů.

Podle zprávy analytika společnosti Andrese Bitterera se každým rokem zdvojnásobuje podíl open source softwaru v oblasti BI, protože dodávané produkty jsou již na dobré úrovni pro běžné aplikace. Prodejci komerčních BI nástrojů mají stále lepší technologie, ale open source se pomalu stává další možností pro každodenní BI řešení. Gartner ve své zprávě uvádí 5 open source organizací: Actuate, Jaspersoft, Jedox, Pentaho, SpagoWorld. Přestože je software ke stažení zdarma, mohou se celkové náklady vyšplhat na tisíce dolarů díky placeným verzím, vysoké vývojové ceně nebo placené podpoře. V porovnání s komerčními řešeními je to však pouze zlomek ceny.[8]

4.1 Magické čtverce – Magic Quadrant



Obrázek 6: Magic Quadrant 2011

Hodnocení Magic Quadrant graficky znázorňuje situaci na trhu k určitému datu a pro určité časové období. Zobrazuje výsledky analýzy provedené společností Gartner, jak jednotliví prodejci splňují kritéria stanovená firmou Gartner pro dané odvětví trhu. Společnost Gartner nepodporuje žádného dodavatele, produkt či službu, které jsou v hodnocení Magic Quadrant uvedeny, ani nedoporučuje uživatelům technologií vybírat si pouze dodavatele zařazené do kvadrantu lídrů (Leaders quadrant). Magic Quadrant je výzkumný nástroj, není zamýšlen jako návod k určitému jednání. Společnost Gartner neposkytuje v souvislosti s tímto výzkumem žádné záruky, výslovné ani předpokládané, jako např. záruky týkajících se prodejnosti či vhodnosti k určitému účelu.[9]

4.2 Open source BI podle Gartner v roce 2010

Společnost Gartner uvádí, že uvažovala o zahrnutí open source BI dodavatelů ve zprávě Magic Quadrant, kteří již dosáhli požadovaných příjmů. Ovšem z důvodu nízkého počtu odpovědí v zákaznickém průzkumu nakonec zahrnutí nebyli. Přestože nesplnili požadavky, ukázali se dva dodavatelé jako životaschopní na trhu BI – Pentaho a Jaspersoft. Oba dodavatelé poskytují komplexní BI platformu, která je srovnatelná v mnoha funkčních oblastech tradičních dodavatelů BI platform. Klíčovou strategií obou dodavatelů je navazovat OEM vztahy s nezávislými dodavateli softwaru, kteří chtějí snadno vložit funkcionalitu BI za nízkou cenu do svých aplikací. Jaspersoft a Pentaho umožňují výrobcům softwaru vložit svoje open source BI komponenty, aniž by byli vázáni podmínkami licence GNU General Public License. Jejich model založený na předplatném poskytuje mimořádnou podporu, což se také odrazilo v zákaznickém průzkumu Magic Quadrant. Jaspersoft i Pentaho dosáhli vysokého skóre v oblasti zákaznické podpory – dokonce vyšší než kterýkoliv velký dodavatel dva roky za sebou.[10]

4.3 Open source BI podle Gartner v roce 2011

Jelikož v prováděném průzkumu zvýšili zákazníci zájem o nižší náklady a počet zákazníků na trhu u open source prodejců rostl, společnost Gartner předpokládala, že Jaspersoft a Pentaho splní kritéria pro zařazení do magických čtverců.

4.3.1 Pentaho

Společnost Pentaho nasbírala dostatek odpovědí v zákaznickém průzkumu a již v říjnu 2010 ohlásila, že její příjmy budou stejné, nebo vyšší, než 15 milionů dolarů. Nicméně společnost Pentaho následně informovala Gartner o tom, že celkové příjmy z jeho BI platformy budou nižší, než se očekávalo. To mělo za následek vyřazení z Magic Quadrant. I přes nesplnění všech kritérií stojí Pentaho určitě za zmínku – poskytuje open source BI platformu tvořenou součástmi ETL, OLAP, reporting, dashboarding, ad hoc analýzy a data mining. Tyto součásti jsou spravované z centrálního BI serveru. Nízká cena licence je klíčovým prvkem hodnoty Pentaha. Zatímco open source dodavatelé mají tendence investovat více v napodobení základních funkcí komerčních BI dodavatelů, Pentaho se zaměřuje na oblast investic do budoucnosti. Zejména byl integrován data mining do BI platformy, nebo přidán ad hoc reporting pro IPAD Apple.

Nicméně podle průzkumu se zdají být integrace platformy, funkčnost produktu a podpory pro zákazníky Pentaha obavou. Přestože Pentaho vydělává na jedné platformě spravované z jednoho serveru, tak úložiště a prostředí zůstávají odděleny. Zákazníci Pentaha ohodnotili integraci jeho BI platformy hůře, než výrobce. K vyřešení tohoto problému vydala společnost v březnu 2010 nástroj Agile BI, který zahrnuje jedno prostředí pro dimenzionální modelování, vytváření datového trhu, ETL,

OLAP. Pentaho si také vysloužilo nejnižší celkové skóre v průzkumu s podprůměrnými výsledky ve všech 13 vlastnostech BI produktu, což je známkou toho, že v produktu stále zůstávají funkční mezery. Kromě toho si v průzkumu vysloužilo nejnižší skóre pro snadnost použití jak pro vývojáře, tak pro koncové uživatele. Ti navrhuji, aby byly dodány nástroje, které splňují rostoucí požadavky pro firemní uživatele, a které umožní intuitivně analyzovat jejich data. Pentaho je založeno na modelu předplatného, závisícího na poskytování kvalitní podpory. Nízké skóre ve všech aspektech podpory produktu, včetně odborné úrovně, rychlosti odezvy a řešení jsou obavy zejména proto, že produkt společnosti je také hodnocen pod průměrem v kvalitě produktu. To dále inklinuje k další interakci zákazníka s podporou. Zatímco Pentaho má některé velmi dobré reference, jeho nasazení mají zhruba poloviční velikost než zjištěný průměr v počtu uživatelů a objemu dat. Většina jeho zákazníků nepovažuje Pentaho za svůj standard, za který uvádějí další výrobce.[11]

4.3.2 Jaspersoft

Jaspersoft se objevuje letos v Magic Quadrant poprvé. Nabízí komplexní open source BI platformu. Jaspersoft Enterprise Edition založená na platformě verze 3.7, včetně JasperReports Server (zahrnující správu serveru, ad hoc dotazů, analýzy in-memory a dashboarding), JasperReports Library, JasperReports iReport Designer, Jaspersoft OLAP, Jaspersoft ETL (což je open source ETL od Talend) a Activity Monitoring Console od Talend (která je součástí komerčního vydání). Jaspersoft 4.0 byl vydán 11. ledna 2011 se stejnými komponentami a předělanou prezentační vrstvou. Vzhledem ke skutečnosti, že zákazníci mohou vkládat svůj software, aniž by byl vázán podmínkami licence GNU General Public License, vydělá Jaspersoft více než polovinu obchodem s OEM výrobcí, kteří zahrnují Jaspersoft jako BI komponenty ve svých softwarových balíčcích, stejně jako jiné společnosti, které integrují Jaspersoft do svých interních aplikací. Vlastní také partnerskou síť, do níž patří společnosti jako VMware, HP, Red Hat a Tata.

Požizovací cena je zdaleka nejpůsobivější část hodnoty Jaspersoftu a je hlavním příčinou jeho úspěchu. Zákazníci uvádějí celkové náklady na vlastnictví jako důvod č. 1 pro volbu Jaspersoftu jako jejich dodavatele BI softwaru, v porovnání s ostatními výrobci je tento důvod uveden nejčastěji, s výjimkou Microsoftu. Dále je příznivá celková cena vlastnictví BI platformy a podprůměrné náklady na implementaci a údržbu na uživatele. Jeho nízkonákladový model se také dobře hodí pro využití v extranetu, kde je často neznámý počet uživatelů. Z průzkumu vyplývá, že je BI platforma od Jaspersoftu používána externě častěji, než platformy od ostatních výrobců.

Zatímco Jaspersoft poskytuje především open source verzi běžných funkcí BI za velmi nízkou cenu licence, jeho roadmapa obsahuje některé prvky, o kterých nepřemýšlí mnoho z jejich komerčních konkurentů. Mezi ně patří vytvoření konektorů pro R-based, revoluční analýzy pro prediktivní analyzování a eXo pro spolupráci s nově vznikajícími zdroji, jako jsou Hadoop, Hive a NoSQL.

Dalším plusem je tzv. „cloud“ nasazení. V roce 2010 spustila společnost Jaspersoft Live, což je zkušební verze jejich BI platformy, ke které může vzdáleně přistoupit každý uživatel přes webové rozhraní. Vyšší procento zákazníků než u ostatních výrobců uvádí, že využívají nebo plánují využít Jaspersoft tímto způsobem.

Upozornění

Jaspersoft tvrdí, že je lídrem v open source BI s více než 13500 komerčních zákazníků, což potvrzují i prokázané tržby. Sekává se však pouze s 20 referenčními zákazníky zahrnující kritéria. To by mohlo být částečně kvůli nedostatku standardních postupů pro správu účtu, jejichž prostřednictvím jsou obvykle zákaznické reference vyvíjeny, a to zejména počet OEM zákazníků, protože ti jsou z účasti v průzkumu vyloučeni.

Jaspersoft inklinuje k nasazení v menších společnostech s menšími počty uživatelů a objemy dat, než je průměr v průzkumu, stejně jako nasazení pouze v některých odděleních oproti celým podnikům. Více než polovina zákazníků považuje jejich BI za standard.

Jaspersoft je primárně nasazován pro reporting, což je hlavní oblast, ve které vyniká. Kvůli tomu dostává nejnížší skóre ze všech dodavatelů, pokud jde o šíři využití funkcionality. Vedle SAPu získal Jaspersoft nejnížší celkové skóre ve funkcionalitě produktu. Kombinace těchto výsledků naznačují, že se Jaspersoft používá v organizacích většinou pro reporting a nejvíce pro lehké ad hoc analýzy. To odpovídá neoficiálním výsledkům průzkumu společnosti Gartner, která naznačuje, že organizace stále více využívají nízkonákladové alternativy, jako jsou open source produkty, pro využití základní funkcionality, jako je reporting, ke snížení celkových nákladů BI portfolia.

Zvláště znepokojivé je, že Jaspersoft dosahuje téměř nejnížšího skóre v oblasti podpory, přestože má obchodní model založený na předplatném. To je významná změna oproti minulému roku, kdy dosahoval Jaspersoft v oblasti podpory vysokého skóre, i když při menším počtu vzorků.

Jaspersoft získal podprůměrné skóre v oblasti snadného použití jak pro koncové uživatele, tak pro vývojáře, podprůměrně byl ohodnocen v integraci jeho BI komponent a nejnížší skóre získal u kvality výrobku, výkonu a prodejní zkušenosti. Přes tyto nevýrazné výsledky spokojenosti zákazníků u nich udržuje do budoucna Jaspersoft pozitivní pohled jako úspěšný dodavatel. Jedno vysvětlení pro tento paradox je, že hodnota nákladů při nasazení Jaspersoftu je nižší, což je v souladu s očekáváním. Problémem je hrozba ze strany Microsoftu. Většina společností, které chtějí nízkonákladový BI software a vlastní již nějaké komponenty od Microsoftu, nebo jsou na Microsoft zaměřeny, si raději pořídí i komponenty Microsoft BI, než open source software. Toto dělá open source BI, včetně Jaspersoftu více relevantní pro společnosti, které nepoužívají software od Microsoftu a také pro OEM partnery.[11]

5 Praktická část

Vzhledem k výsledkům dosažených v první části práce jsem se rozhodl otestovat nástroje Pentaho a Jaspersoft. Práci jsem rozdělil na několik částí:

1. Stažení a instalace programů
2. Testování funkcionality nástrojů
3. Zátěžový test
4. Hodnocení

5.1 Stažení a instalace

Pro stažení nástrojů Pentaho i Jaspersoft je nutné se nejprve zaregistrovat na jejich webových stránkách, poté máme programy volně k dispozici. U obou produktů se nabízí dvě možnosti, a to volná komerční verze na 30 dní v podobě jednoho instalačního souboru, nebo komunitní verze, u které si stáhneme několik archivů pro jednotlivé nástroje. Samotná instalace proběhla bez větších problémů.

5.2 Testování funkcionality nástrojů

5.2.1 Pentaho

Po úspěšné instalaci můžeme začít pracovat s nástrojem Pentaho BI suite. Práce probíhá pře webový prohlížeč a to standardně na portu 8080, k dispozici je také administrativní konzole na portu 8088, kde je možno spravovat uživatele, datové zdroje, či nastavovat různé parametry. Při instalaci byl také vytvořen datový zdroj s ukázkovými daty, nad kterými je možno ihned po spuštění vytvářet různé reporty či analýzy.

5.2.1.1 Webová aplikace

Pro přihlášení je nutné zadat uživatelské jméno a heslo, po instalaci je vytvořeno několik uživatelů automaticky, takže stačí vyplnit některého z těchto uživatelů. Po úspěšném přihlášení se spustí samotná aplikace. V horní části jsou umístěny ikony pro tvorby nových reportů, analýz a dashboardů. V levé části je umístěno rozbalovací menu, ve kterém můžeme procházet jednotlivé adresáře, ve kterých jsou umístěny všechny pracovní soubory, ať už reporty nebo analýzy či grafy, které je možno vidět v levé spodní části. V pravé části obrazovky je umístěna pracovní plocha. První položkou v rozbalovacím menu jsou analýzy. Po kliknutí se nám ve spodní části zobrazí již vytvořené soubory, a tak po otevření některého z nich můžeme vidět hotovou analýzu. Ta se obvykle skládá z tabulky, ve které jsou uvedena data a je možné ji různým způsobem rozbalovat či upravovat a dále pak grafického vyjádření těchto dat v podobě grafů. Pro práci s analýzami máme k dispozici novou nabídku 17 různých ikon. Ty slouží pro úpravu OLAP kostky, MDX dotazy, konfigurace OLAP tabulky, přepínání mezi osami, konfiguraci grafu, zapnutí či vypnutí grafu, nastavení pro tisk či okamžitý export do excelu. Vytvoření nové analýzy je poměrně snadné. Stačí si vybrat schéma a kostku, nad kterou chceme analýzu vytvořit a po té například upravit MDX dotaz, aby se nám zobrazovala požadovaná data.

Druhou položkou v menu jsou grafy, kde je opět na ukázkou vytvořeno několik jednoduchých grafů. Zajímavější je ovšem položka reportů. Můžeme si spustit již vytvořené reporty, jako v předchozích případech, nebo se rovnou pustit do vytváření vlastního reportu. To se skládá ze 4 kroků. Nejdříve je nutné vybrat správný datový zdroj a také zvolit vzhled reportu, dále budeme vybírat položky (sloupce) z datového zdroje, které zde chceme mít zobrazeny, ty je možné v dalším kroku

ještě poupravit a v posledním kroku zbývá už jen nastavení celého reportu, jako například záhlaví, zápatí či popis. V těchto 4 krocích jsme schopni snadno vytvořit jednoduché reporty v různých formátech. Další položkou v menu jsou dashboards. Jedná se o interaktivní panely, které jsou schopny reagovat na změnu parametru a okamžitě zobrazovat výsledky. Můžeme si spustit kterýkoliv z vytvořených a podívat se, jak fungují.

K dispozici máme dále stejné nástroje, které si u komunitní edice stahujeme odděleně. Můžeme si tak spustit nástroje jako Report Designer, Data integration či Metadata Editor.

5.2.1.2 Pentaho Report Designer

Report Designer je desktopová aplikace, která poskytuje vizuálně-návrhové prostředí pro vytvoření definice reportů. Tyto definice mohou být prováděny a ukládány lokálně přes Pentaho Report Designer nebo publikovány na Pentaho BI serveru, který umožňuje mnoha uživatelům přistupovat a plánovat vykonávání reportů. Report Designer je zaměřen na zkušené a pokročilé uživatele, kteří jsou obeznámeni s pojmy a použitím datových zdrojů.

5.2.1.3 Tvorba reportů

Vytvoření základních reportů není nijak náročné, po spuštění programu máme možnost začít pracovat s prázdným listem a manuálně si připravit report podle svých představ, nebo zvolit druhou variantu a pracovat podle průvodce. Podobně jako ve webové aplikaci, i zde report vytvoříme ve 4 krocích. Nejprve je nutné si zvolit vzhled reportu, na výběr máme několik připravených vzorů. V druhém kroku určíme datový zdroj, ze kterého budeme čerpat data, po té je nutné vybrat sloupce, které budeme v reportu zobrazovat a nakonec můžeme přidat k některým prvkům popisky. Po dokončení průvodce dostaneme výsledný report na pracovní plochu, kde s ním můžeme dále pracovat a přidávat různé prvky, jako jsou obrázky, grafy, popisky a další, či upravovat ty současné. Pokud je práce hotová, stačí report uložit, exportovat do formátu pdf, rtf, html, csv a dalších, nebo přímo publikovat na BI server.

5.2.1.4 Vytvoření datového zdroje

Pro vytváření datových zdrojů slouží ikona žlutého válce. Nabízí se několik různých možností datového zdroje: JDBC, Pentaho Data Integration, OLAP, Metadata, XML, Tabulka či pokročilejší volby, jako jsou skripty a java. Jelikož jsem chtěl otestovat, jak program funguje s přístupem do různých databází, pracoval jsem s JDBC konektorem. Datový zdroj vytvoříme v několika krocích. Buď můžeme upravovat již vytvořené datové zdroje, nebo přidáme nové. Nejdůležitější je výběr databáze, ke které se chceme připojit. Na výběr máme téměř 40 různých databází, jako například Oracle, MySQL, PostgreSQL, Netezza, Teradata, Hypersonic a mnoho dalších. Po výběru databáze je nutno vyplnit umístění databáze, její jméno, číslo portu a přihlašovací údaje. Takto vyplněné připojení otestujeme, a pokud je vše v pořádku, máme vytvořený funkční datový zdroj, nad kterým můžeme vytvářet reporty.

Vytvořil jsem připojení do databází Hypersonic, Oracle, MySQL a načítání dat probíhalo v pořádku. Databáze Hypersonic je nainstalovaná spolu s programem Report Designer, u Oracle nenastal s JDBC ovladačem žádný problém a práce probíhala taktéž bezchybně, pro připojení MySQL jsem nainstaloval na lokálním počítači MySQL server a přistupoval k němu, i poslední možnost byla bez problémů.

5.2.1.5 Pentaho Data Integration (Kettle)

Kettle poskytuje výkonné ETL funkce pomocí inovativního přístupu řízením metadat. S intuitivním, grafickým, drag and drop prostředím a osvědčenou, škálovatelnou, standardně postavenou architekturou, je Pentaho Data Integration dobrou volbou pro organizace, ve srovnání s tradičními, proprietárními ETL nástroji.

Nástroj se skládá z několika částí:

Spoon – jedná se o grafický nástroj určený pro modelování datových toků od vstupu až po výstup. Každý takový model se nazývá transformace.

Pan – tento nástroj provádí pomocí příkazové řádky transformace vymodelované ve Spoon.

Kitchen – je určený pro vykonávání tzv. jobů z příkazové řádky. Pod pojmem Job si můžeme představit například pravidelné spouštění transformací, správu souborů či zasílání emailových zpráv.

5.2.1.6 Pentaho Data Mining (WEKA)

WEKA je komplexní sada nástrojů pro dolování dat. Jeho široká nabídka funkcí jako klasifikace, regrese, asociační pravidla, shlukovací algoritmy mohou být použity, aby pomohli lépe pochopit podnikání a také být využity ke zlepšení budoucí výkonnosti prostřednictvím prediktivní analitiky.

5.2.1.7 Rozdíly mezi edicemi

Pentaho je nabízeno ve dvou edicích, a to komunitní open source edici a komerční enterprise edici. Zatímco u komunitní verze jsou jednotlivé nástroje licencovány samostatně a jsou volně dostupné jako open source, u enterprise edice jsou nabízeny pod komerční licencí a jsou certifikovány. Velkým rozdílem mezi těmito verzemi je nabízená podpora. U komunitní edice máme možnost komunikovat na fóru, kde není nikdy odpověď garantovaná, u enterprise edice je však dostupná profesionální podpora, která je připravena řešit jakékoliv problémy. Dalším rozdílem mezi oběma edicemi je také funkcionality. Placená verze nabízí navíc například tyto funkce:

Pentaho analyzer, podmíněné formátování, enterprise konzole, efektivní konfigurace zabezpečení, aplikační diagnostika, skladové utility, management životního cyklu, audit reporty, automatizované vypršení platnosti obsahu, clustering, monitoring výkonu, ETL řízení a monitorování, integrace s třetí stranou LDAP / MSAD, zabezpečení oprávnění obsahu, rozšířená bezpečnost úložiště, rozšířená správa obsahu, vzdálené plánování úloh a správy, úložiště týmového vývoje včetně verzování a zamykání, a další. Podle průzkumu, který provádí společnost Gartner, se pohybuje průměrná cena ročního předplatného u Pentaho Enterprise Edice kolem 24 000 USD.

5.2.2 Jaspersoft

Instalace proběhla bez problémů a tak bylo možné začít pracovat balíkem Jaspersoft BI suite. Práce probíhá také přes webový prohlížeč a po startu BI serveru se nám zobrazí úvodní přihlašovací obrazovka. Pro přihlášení použijeme přednastavené údaje a potvrdíme. Úvodní obrazovka je rozdělena na dvě části, kde vlevo je seznam již vytvořených sestav či grafů, které se následně po kliknutí zobrazují v druhé části vpravo. K dispozici máme také menu v horní části, pomocí kterého se celá aplikace ovládá. Vlevo je ikonka domu, která nás odkazuje na úvodní stránku, druhou položkou v menu je „View“, která po rozbalení nabízí 5 možností: „Search results“ – slouží pro vyhledávání. „Repository“ – pro zobrazení repositářů. „Reports“ – pro zobrazení vytvořených reportů. „OLAP views“ – pro zobrazení vytvořených OLAP pohledů. U každé OLAP tabulky máme potom k dispozici několik ikon pro její konfiguraci, jejich funkcionality je téměř stejná jako u Pentaha. Poslední v této části menu je položka „Messages“, pod kterou se zobrazují zprávy a vzkazy. Třetí a poslední je v hlavním menu položka „Create“. Obsahuje „Ad hoc report“ pro tvorbu reportů a „Dashboard“ pro tvorbu dashboardů.

Vytvoření reportu je poměrně jednoduché, stačí si vybrat datový zdroj a pomocí průvodce si vybrat, co všechno chceme mít v reportu zobrazené. V dalším kroku je možné si přetáhnout libovolné sloupce na pracovní plochu a dle libosti je seřadit či upravovat vzhled reportu. Pokud máme práci hotovou, můžeme report uložit a následně si ho zobrazit a vyexportovat například do formátu pdf, rtf, docx, xlsx, nebo jiných.

U tvorby dashboardu jsou v levém menu různé prvky, které si můžeme jednoduše přetáhnout do hlavní pracovní plochy, jedná se například o jednoduchý text, tlačítka, grafy nebo i kompletní reporty. Po vytvoření dashboard uložíme a můžeme si ho opět zobrazit.

5.2.2.1 Jaspersoft iReport Designer

Podobně jako u Pentany, máme k dispozici i u Jaspersoftu jednotlivé nástroje, pro reporting je to iReport Designer. Po spuštění programu se můžeme rozhodnout, jestli budeme report tvořit kompletně od začátku, nebo použijeme některou z předem připravených šablon. Následně je nutné report pojmenovat a zvolit umístění, ve třetím kroku musíme nastavit datový zdroj, ze kterého chceme čerpat data. Poté už jen vybereme ty sloupce, které chceme mít v reportu zobrazeny a uspořádáme si je tak, jak chceme. Pak stačí potvrdit dokončení a report je hotov. Výsledný report můžeme na pracovní ploše ještě upravovat a nakonec uložit, nebo exportovat do různých formátů (pdf, html, doc a další). Postup tvorby reportů je téměř stejný jako u Report Designeru od Pentaha a co se týče možností, jsou tyto programy na podobné úrovni.

I postup tvorby připojení k datovému zdroji je velice podobný, po kliknutí na ikonu datového zdroje se otevře okno nastavení, kde můžeme upravovat stávající, nebo vytvořit nové připojení. Jaspersoft nabízí u typu datového zdroje o několik možností více, jako například XMLA server, EJBQL či Hibernate. I zde jsem testoval připojení pomocí JDBC do několika databází (Oracle, MySQL, HSQLDB). V nabídce databází je iReport o něco chudší a k dispozici máme cca polovinu z nabídky Pentaha. Problém nastal pouze v případě připojení k databázi Oracle, kdy bylo nutné nakopírovat JDBC ovladač do složky k dalším ovladačům. V ostatních případech pracoval nástroj správně.

5.2.2.2 JasperETL

JasperETL je postaven nad nástrojem od společnosti Talend. Je to velice flexibilní, výkonný a cenově dostupný open source nástroj pro oblast integrace dat. Je určen na podporu jednoho, nebo více vývojářů při škálování na nejvyšší úrovni. Vývojáři mohou graficky navrhovat, plánovat a provádět pohyby a transformace dat pro Business Intelligence projekty, jako je například nahrávání do Operational Data Store, Datamartů, nebo datových skladů. JasperETL umožňuje organizacím vyvíjet integrační procesy a spravovat data pro lepší reporting a analýzy.

5.2.2.3 Rozdíly mezi edicemi

Jaspersoft je rozdělen na Community projects (CP), Professional Edici (PE), Enterprise Edici (EE). CP poskytuje BI funkce pro open source vývojáře a technologické nadšence, kteří chtějí přidat reporty a analýzy do jejich provozních procesů. Software může být upraven a distribuován v souladu s podmínkami GNU GPL. Professional Edice je určena pro organizace, které vyžadují pokročilé funkce s jistotou profesionální kvality technické podpory a komerční licenční ochranu.

Enterprise Edice je postavena pro nejnáročnější požadavky na BI, včetně podpory SaaS, integrace dat a OLAP analýzy. Enterprise Edition přichází také s profesionální kvalitou technické podpory a komerční licenční ochranou. Kromě těchto edicí, mohou zákazníci případně zakoupit jednotlivé produkty, aby vyhovovaly jejich vlastním požadavkům OEM licencování. Professional Edice je k dispozici také v rámci OEM licence. S licencí OEM dostanete stejný produkt, podporu a obchodní výhody, plus práva vložit nebo redistribuovat Jaspersoft v komerční distribuci On Demand nebo Software jako služba (SaaS).

Enterprise Edice nabízí na rozdíl od Community Projects také: Flash mapy, rozšířené grafy, Ad hoc dotazování a reporting, dashboardy, SaaS, audity, aktualizace, profesionální podporu, profesionální dokumentaci, modifikovatelný zdrojový kód, podporu certifikované platformy, vkládání komerčních práv.

5.3 Zátěžový test – ETL nástroje

Abych prakticky otestoval, jakého výkonu dosáhnou nástroje pro datovou integraci – tedy Pentaho Data Integration (Kettle) od společnosti Pentaho a JasperETL od Jaspersoftu, vytvořil jsem několik datových transformací shodných pro oba nástroje. Testy byly provedeny nad vygenerovanými daty, jelikož jsem nenašel žádný zdroj, který by mi poskytl potřebné množství dat. V těchto testech nebyla hlavní kvalita dat, ale spíše jejich množství, na kterém bylo možno otestovat výkonnost a rychlost obou nástrojů. Každá z 8 transformací byla provedena 3x nad daty o velikosti 100 000, 1 000 000 a 5 000 000 řádků, což je celkem 9 výsledků a z nich byl následně vybrán nejlepší výsledek pro každou transformaci.

Testy byly prováděny na PC s procesorem Intel Core 2 o frekvenci 2,13GHz, který měl operační paměť 2GB RAM a operační systém Window XP Professional Service Pack 3.

V příloze jsou popsány jednotlivé transformace, u každé z nich je uvedeno schéma z obou nástrojů, dále tabulka výsledků měření a grafické znázornění těchto hodnot. Z celkových 8 testovaných transformací dosáhlo Pentaho Data Integration (Kettle) lepších výsledků pouze jednou – a to v případě čtení z databáze MySQL a následného zápisu do souboru XML. Ve všech ostatních testech měl lepší výsledky Jaspersoft, a to i několikanásobně. Zatímco ve většině případů se rychlost transformací pohybovala u PDI (Kettle) cca kolem 30 000 řádků za sekundu, u JasperETL to bylo často přes 100 000 řádků za sekundu, v jednom případě dokonce překonal hranici 200 000 řádků. V tomto zátěžovém testu u mě zvítězil určitě JasperETL.

5.4 Hodnocení

Práci s oběma nástroji hodnotím kladně, po seznámení se s nimi pracovalo dobře, a zručnost v jejich ovládání se s postupem času zlepšovala. Prostředí nástrojů nebyla nijak složitá, většina práce proběhla intuitivně, bez vážnějších problémů. Díky mnoha dostupným příručkám od obou výrobců nebyl většinou problém vyhledat potřebné informace. Co se týká funkcionality nástrojů, Jaspersoft i Pentaho jsou na tom srovnatelně a poskytují dostatečné množství funkcí, aby mohly být nasazeny do malých i středních společností. Vzhledem k tomu, že je jejich vývoj řízen podle roadmap a snaží se dohnat komerční výrobce, lze přepokládat, že se i nadále bude zvyšovat podíl open source řešení na trhu.

6 Závěr

Moje bakalářská práce byla vypracována v rámci odborné praxe ve firmě Tieto Czech s.r.o. v oddělení Business Intelligence. Hlavním důvodem zpracování tohoto tématu bylo, aby byla firma schopná zareagovat na případné požadavky nízkonákladových BI řešení a měla přehled o tom, co může zákazníkům nabídnout.

Celá práce byla rozdělena na několik částí. V první fázi jsem získával teoretické znalosti o Business Intelligence, a to z různých firemních prezentací o BI, z internetu či dostupné literatury, abych je pak dále v práci mohl uplatnit.

V druhé fázi bylo nutné zjistit, jaké jsou komplexní open source nástroje dostupné v oblasti BI. Po analýze trhu jsem se zaměřil na následující nástroje: Pentaho, Jaspersoft, Spago, Palo a OpenI. Pro jejich srovnání jsem předem zvolil určitá kritéria a následně jsem produkty podle nich ohodnotil. Vítězí mého hodnocení se stalo Pentaho, které následoval Jaspersoft s menší ztrátou, tyto dva produkty jsem se pak srovnával i po praktické stránce v poslední části.

Fakt, že Pentaho a Jaspersoft dosáhli nejvyššího hodnocení v mém porovnání je určitě dobrým výsledkem, protože tyto nástroje se objevují také ve zprávách společnosti Gartner, která provádí nezávislé průzkumy v oblasti informačních technologií. Splněním mnoha kritérií dokazují tito dodavatelé, že mohou konkurovat i komerčním produktům. Podíl open source nástrojů se na poli BI každoročně zdvojnásobuje.

Společnosti, které si nemohou dovolit drahé řešení, hledají alternativní možnosti. Open source nástroje mají při správné volbě a nasazení co nabídnout a do určité míry dokážou nahradit komerční produkty. Hlavní výhodou jsou nízké počáteční náklady a možnost úpravy softwaru podle vlastních potřeb. Dodavatelé open source produktů mají hlavní zisky z poskytované podpory či různých školení. Zatímco u placené edice je profesionální podpora garantovaná, u bezplatných verzí budeme odkázáni na dokumentace či fóra.

V poslední fázi jsem nástroje od Pentaha a Jaspersoftu prakticky testoval. Během práce s nimi nevznikly žádné závažnější problémy. Uživatelské prostředí bylo přirozené a práce probíhala logicky a poměrně jednoduše. Funkcionalita obou produktů je srovnatelná. Programy pracovaly bez potíží a splnily očekávání.

Ve srovnání v zátěžovém testu nástrojů pro ETL dosáhl lepších výsledků Jaspersoft. Produkt od této společnosti ovšem na druhou stranu neobsahuje nástroj pro data mining. Při volbě BI řešení je tedy nutné dobře zvážit, co od produktu očekáváme. Jaspersoft se dříve zaměřoval hlavně na reporting, což byl hlavní důvod, proč si ho zákazníci pořizovali. Dnes už nabízí i další nástroje pro analýzy a ETL.

Pentaho BI suite obsahuje také nástroj Weka pro data mining. Lze tedy říci, že open source nástroje od společností Pentaho a Jaspersoft mohou být nasazeny u malých a středních firem jako nízkonákladové BI řešení.

7 Literatura

- [1.] *Business Intelligence – Wikipedie, otevřená encyklopedie*. [Online] 12. 4. 2011. [Citace: 15. 4 2011.] <http://cs.wikipedia.org/wiki/Business_Intelligence>.
- [2.] NÁVRAT, Radan. Business Intelligence jako open source. *IT Systems*. [Online]. 2009, roč. 10, č. 5 [Citace: 16.3.2011.] <<http://www.systemonline.cz/Business-Intelligence/Business-Intelligence-jako-open-source.htm>>.
- [3.] PŘIBYSLAVSKÝ, Jiří. Business Intelligence – opřete se o reálná čísla. *IT Systems*. [Online]. 2010, roč. 11, č. 10 [Citace: 16.3.2011.] <<http://www.systemonline.cz/Business-Intelligence/Business-Intelligence-oprete-se-o-realna-cisla.htm>>.
- [4.] SCHILLER, Martin. Co se skrývá pod zkratkou ETL ?. *IT Systems*. [Online]. 2003, roč. 4, č. 3 [Citace: 16.3.2011.] <<http://www.systemonline.cz/clanky/co-se-skryva-pod-zkratkou-etl.htm>>.
- [5.] *OLAP – Wikipedie, otevřená encyklopedie*. [Online] 14.12.2010. [Citace: 15. 3 2011.] <<http://cs.wikipedia.org/wiki/OLAP>>.
- [6.] *Data Mining – Wikipedie, otevřená encyklopedie*. [Online] 21. 4. 2011. [Citace: 22. 4 2011.] <http://cs.wikipedia.org/wiki/Data_mining>.
- [7.] *SaaS– Wikipedie, otevřená encyklopedie*. [Online] 8.3. 2011. [Citace: 22. 4 2011.] <<http://cs.wikipedia.org/wiki/SaaS>>.
- [8.] BETTS, Mitch. Gartner backs open source BI tools. *Computerworld*. [Online]. 5.1.2010. [Citace: 16.3.2011.] <<http://news.techworld.com/applications/3209484/gartner-backs-open-source-bi-tools>>.
- [9.] *Podle hodnocení Gartner patří Qliktech s produktem QlikView mezi leadry v oblasti Business Intelligence – Lupa.cz*. [Online] 14.2 2011. [Citace: 15.3. 2011.] <<http://www.lupa.cz/tiskove-zpravy/podle-hodnoceni-gartner-patri-qliktech-s-produktem-qlikview-mezi-leadry-v-oblasti-Business-Intelligence>>.
- [10.] SALLAM, Rita L. – HOSTMANN, Bill – RICHARDSON, James – BITTERER, Andreas. Magic Quadrant for Business Intelligence Platforms. *Gartner*. [Online]. 29.1.2010. [Citace: 16.3.2011.] <<http://www.gartner.com/technology/media-products/reprints/oracle/article121/article121.html>>.
- [11.] SALLAM, Rita L. – RICHARDSON, James – HAGERTY, Jonh – HOSTMANN, Bill. Magic Quadrant for Business Intelligence Platforms. *Gartner*. [Online]. 27.1.2011. [Citace: 16.3.2011.] <<http://www.gartner.com/technology/media-products/reprints/qliktech/article3/article3.html>>.
- [12.] NOVOTNÝ, Ota – POUR, Jan – SLÁNSKÝ, David. *Business Intelligence. Jak využít bohatství ve vašich datech*. 1. vyd. Praha: Grada Publishing, 2004. 256 s. ISBN 80-247-1094-3.

Seznam příloh

Příloha č. 1: Zátěžový test ETL nástrojů

Adresářová struktura přiloženého DVD

/Prilohy	Adresář obsahuje přílohy k práci
/Prirucky	Stažené příručky k nástrojům
/Texty	Soubory s textem práce, zadání, klíčová slova a abstrakt